**CS 134: Networks** Lecture 5, Monday, 2/6

Prof. Yaron Singer Spring 2017

# 1 Overview

In the past two lectures, we have been interested in mathematical models of social networks that can describe the small-world phenomenon. In this lecture, we will be interested in another characteristic of social networks: their *degree distribution*. Equivalently, we wish to answer the following question:

*As a function of $d$, what is the fraction of nodes in the network which have degree $d$?*

The above question can be thought in terms of popularity in a social network: how many persons have $d$ friends?

# 2 Normal Distributions

Before we answer this question, it is worth considering the normal distribution (see figure below). Many phenomena in nature have behavior that can be well described in terms of a normal distribution, and this seems like a reasonable guess for degree distributions in social networks.

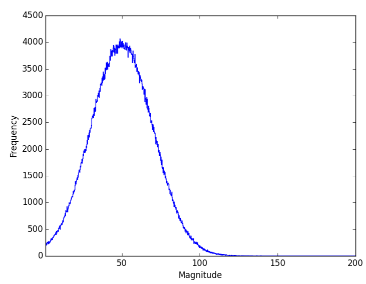


Figure 1: Density

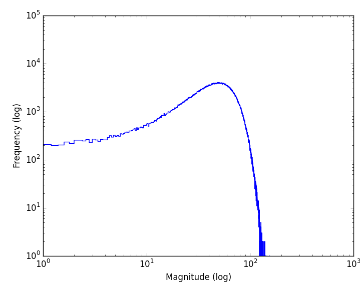


Figure 2: Density (log-log scale)

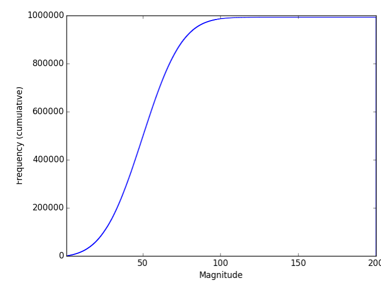


Figure 3: CDF

In the early days of the web, researchers analyzed the web graph and to their surprise discovered that the degree distribution does not at all look like a normal distribution. Instead, they saw that the fraction of webpages with degree $d$ was approximately proportional to $1/d^2$. This is what we call a *power law* distribution.

# 3 Power Law Distributions

Social networks, much like the web graph, tend to have a *power law* degree distribution. Power law degree distributions are characterized by having the number of nodes of degree $d$ to be proportional to $d^{-\alpha}$ for some constant $\alpha$.

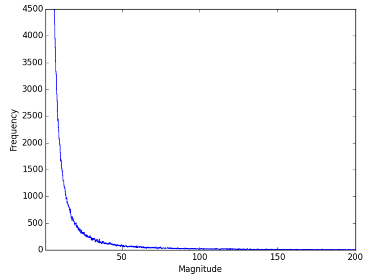


Figure 4: Density

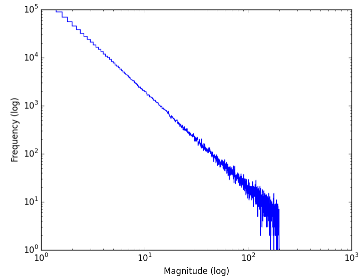


Figure 5: Density (log-log scale)

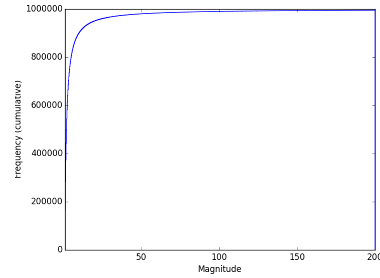


Figure 6: CDF

**Degree distribution as a random variable.** Recall that the degree distribution of a graph is simply a table that summarizes how many nodes there are of a particular degree. When we are given a fixed graph, we can write a simple script and generate this table to get this basic statistic. Although there is no randomness involved in a fixed graph, it will be convenient to think about the degree of a node as a random variable. To do so, we assign $X$ to be the random variable associated with the degree of a node when selecting nodes uniformly at random from the graph.

**Definition.** *A non-negative random variable $X$ has a power law distribution if there exist constants $c, \alpha > 0$ such that:*

$$Pr[X \geq x] = c \cdot x^{-\alpha};$$

*In the case where $X$ is a discrete random variable, then there exist constants $c, \alpha > 0$ s.t.:*

$$Pr[X = x] = c \cdot x^{-\alpha}.$$

Remember that it is important for a given distribution that it has valid parameters. Recall that a discrete nonnegative random variable $X$ with a sample space $\mathcal{S}$ must satisfy $\sum\limits_{x=0}^{\infty} \Pr(X = x) = 1$. Similarly, a continuous nonnegative random variable $X$ with a density function $p(x)$ must satisfy $\int_0^\infty p(x)dx = 1$. For power law distributions, there are two important considerations with regards to ensuring a valid distribution:[1]

- It is impossible for a random variable $X$ that follows a power law distribution with parameters $c, \alpha$ to take on a value of zero; this is true because $cx^{-\alpha}$ is undefined at $x = 0$, as $0^{-\alpha} = 1/0$ is undefined for all $\alpha > 0$. Thus, a minimum value $x_{min} > 0$ for the support of $X$ is typically provided, and we pick the normalizing constant $c$ such that $\sum\limits_{x=0}^{\infty} \Pr(X = x) = 1$, where $\Pr(X = x)$ is defined by

$$\Pr(X = x) = \begin{cases} 0 & \text{if } x < x_{min} \\ cx^{-\alpha} & \text{otherwise} \end{cases}.$$

- Further, if $0 < \alpha \leq 1$, the sum $\sum\limits_{x=x_{min}}^{\infty} cx^{-\alpha}$ diverges; thus, a maximum value $x_{max}$ must be

[1] Note that I while I use notation for the discrete case, the same notions apply to the continuous case. Also, for more related discussion, see the Section 4 notes, especially on the Normalizing Factor $c$.

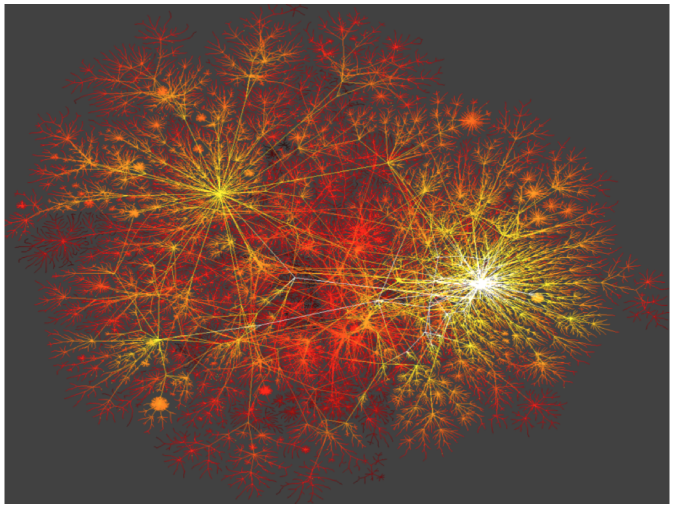

Figure 7: Illustration of a social network where nodes are colored as a function of their degree.

provided in this case[2]. Then, we pick the normalizing constant $c$ such that $\sum_{x=0}^{\infty} \Pr(X = x) = 1$ where $\Pr(X = x)$ is defined by

$$\Pr(X = x) = \begin{cases} 0 & \text{if } x < x_{min} \text{ or } x > x_{max} \\ cx^{-\alpha} & \text{otherwise} \end{cases}.$$

### 3.1 Properties of Power Laws

**Linearity on log-log plot.** When a power law degree distribution is plotted on a logarithmic scale for both the horizontal and vertical axis, we observe a straight line. This behavior is explained mathematically. Let $f(d)$ be the fraction of nodes in a network that have degree $d$. Recall that since we are in the discrete case, we have that $f(d) = c \cdot d^{-\alpha}$ for constants $c, \alpha > 0$. Taking the logs, we get $\log f(d) = -\alpha \log(d) + \log(c)$. Therefore, on a log-log plot, we obtain a straight line with slope $-\alpha$.

**Heavy tail.** An important property of power law degree distributions is their *heavy tail*: they decrease at a polynomial speed rather than exponential. Figure **??** shows a degree distribution that exhibits a power law behavior.

The behavior of a power law degree distribution

$$\Pr(X \geq x) = c \cdot x^{-\alpha}$$

is said to have *polynomial decay* since the distribution decreases polynomially in $x$ as $x$ increases. This polynomial decay is in contrast with the *exponential decay* of, for example, the normal distribution with mean $\mu$ and variance $\sigma^2 = 1$ where

$$\Pr(X \geq x) = \int_{y=x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2}} dy.$$

[2]This case will be discussed further in the next lecture.

The consequence of this polynomial decay is that values are less concentrated around the mean and the values that are distant from the mean form a *heavy tail*.

### 3.2 History of Power Laws

Power laws have been studied since long before the explosion of social networks. They have been empirically observed and studied in many different contexts. Two such examples are in city populations (Figure 10) and income distribution (Figure 11).

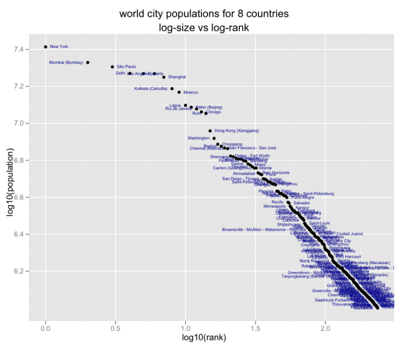


Figure 8: World city populations for 8 countries.

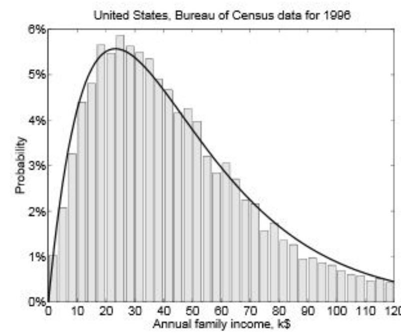


Figure 9: Income distribution in the United States.

## 4 Justification of Power Law in Language

Suppose that we want to design a language that optimizes the average amount of information per unit transmission.

- A language has $n$ words and $d$ letters.
- The $j$th most used word is used with probability $p_j$.
- The costs of sending the $j$th most used word is $c_j = \log_d j$.

The average cost $C$ of transmission is then

$$C = \sum_{j=1}^{n} p_j c_j = \sum_{j=1}^{n} p_j \log_d j.$$

The average information per word is measured using the entropy $H$,

$$H = -\sum_{i=1}^{n} p_j \log_2 p_j.$$

To optimize the average amount of information per unit transmission cost, we wish to pick the probabilities $p_j$ that maximize

$$A := \frac{H}{C}.$$

By taking the derivative, we get

$$\frac{\partial A}{\partial p_j} = \frac{c_j \cdot H + C \cdot \log_2(e \cdot p_j)}{H^2},$$

which is equal to zero, and therefore optimized, at

$$p_j = \frac{2^{-H \cdot c_j / C}}{e}.$$

Since $c_j = \log_d j$, we obtain that the optimal $p_j$s follow a power law.